ownCloud®

Your Cloud, Your Data, Your Way

The File Sync Algorithm of the ownCloud Desktop Clients

Workshop on Cloud Services for File Synchronisation and Sharing

November 17-18, 2014, CERN

Klaas Freitag

ownCloud Client Developer

freitag@owncloud.com

Klaas Freitag

- Free software developer by passion
- Started on the client in late 2011
- Changed from Linuxdistributor SUSE to the ownCloud company in February 2012

Topics

- Syncing – General Thoughts
- Three phases of ownCloud Sync Protocol
- Some Details on Decision Taking and Execution
- WebDAV: Weaknesses and Extensions
- Possible Extensions and Improvements

What is Syncing?

- two collections of data is kept in exact the same state: data, metadata
- Changes are triggered equally by both sides
- No time constraints
- Correctness: Never data loss
- Ability to deal with big data sets
- No user interaction

General Thoughts:

ownCloud

- Open source software
- High expectations
- Standard compliance: WebDAV, Authentication etc.

History

- Started in late 2011 as a “Hackweek” project
- From Feb. 2012 Desktop Client Team at ownCloud Inc.
- Csync and mirall

Three Phases of ownCloud Syncing

One “Sync run”

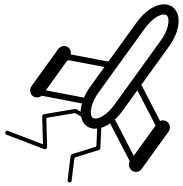
1. Discovery

2. Reconcile

3. Propagation

Three Phases of ownCloud Syncing

A sync run is triggered:



At the program start

Regularly after a certain time

Local repository: On notification

… at some other events

Three Phases of ownCloud Syncing

Discovery phase: Collecting data

Local

ownCloud

Inode

hardlink count

mode

File id

Remote permissions

Sharing state

ETag

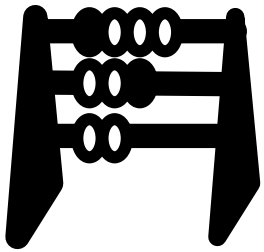
File metadata: Size, modification time, filename (utf-8), phash

Result: Two trees of file meta data in memory.

Reconcile phase: Decision taking

Iterate over both file trees from discovery phase.

Decide what to do with each file based on the collected meta data.



Result: A working list for propagation.

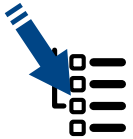
Reconcile phase: Examples

Local file changed:

Modification time **of local file is different to the one in the sync journal**

Remote file changed:

Etag **is different as the one saved in the sync journal**



Local file renamed:

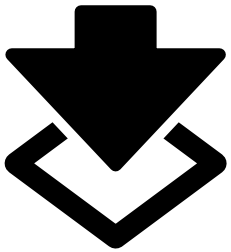
Local file with name is missing. Another file with the same *inode* exists on local repository.

Remote file renamed:

Remote file with name is missing. A file with different name but same file id exists on server.

Propagator phase: Execute decisions

- Execute the action job list per directory.
- Local repository: Platform dependent file system operations
- Remote repository: HTTP/WebDAV operations like PUT, GET, MKCOL, MOVE, PROPSET
- Network operations run in parallel
- Sync journal maintenance, error handling, progress...



Propagator phase: Details

- Big file chunking
- Files change during upload
- Conflict handling

WebDAV Shortcomings

- Performance!
- Everything is a request
- Error reporting
- Session handling

ownCloud WebDAV Extensions

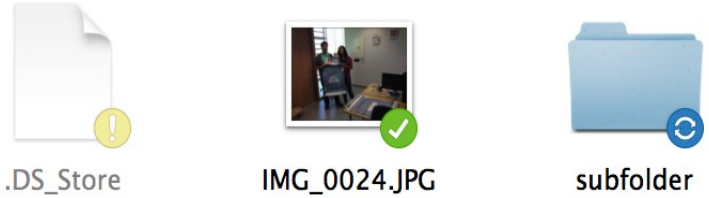
For functionality and efficiency

- File ids to detect remote renames
- Recursive Propfind
- Etag propagation server side
- Additional HTTP headers to transmit meta data

Future: Desktop Integration

- **Previous** Versions: Little OS integration

- Version 1.7.0: Overlay Icons
  - Sync status
  - Sharing state



- Future Versions:
  - Share from desktop
  - Automated upload of attachments to ownCloud Server
  - User interface improvements
    - Tray menu
    - Sync progress display

Example 3: Convenience

- Upcoming Version: Selective Sync
  - Exclude parts of the file tree from the sync to save harddrive space and bandwith
  - No need to reconfigure sync connections
  - ETA calculation in the progress bar
- Future Versions:
  - Admin driven configurations
  - Activity monitor incl. Activities from server